Journal of Biomolecular NMR, **27:** 41–56, 2003.
KLUWER/ESCOM

Determination of molecular alignment tensors without backbone resonance assignment: Aid to rapid analysis of protein-protein interactions

Markus Zweckstetter*
Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany

Received 9 December 2002; Accepted 11 April 2003

Key words: alignment tensor, dipolar coupling, homodimerization, malate synthase G, maltodextrin-binding protein, protein-ligand docking, protein-protein interaction, singular value decomposition, structural genomics

Abstract

Based on high-resolution structures of the free molecules accurate determination of structures of protein complexes by NMR spectroscopy is possible using residual dipolar couplings. In order, however, to be able to apply these methods, protein backbone resonances have to be assigned first. This NMR assignment process is particularly difficult and time consuming for protein sizes above 20 kDa. Here we show that, when NMR resonances belonging to a specific amino acid type are selected either by amino acid specific labeling, by their characteristic C^{α}/C^{β} chemical shifts or by dedicated NMR experiments, molecular alignment tensors of proteins up to 80 kDa can be determined without prior backbone resonance assignment. This offers the opportunity to greatly accelerate determination of three-dimensional structures of protein-protein and protein-ligand complexes, and validation of multimeric states of proteins. Moreover, exhaustive back-calculation can be performed using only $^{1}D_{NH}$ couplings. Therefore, it avoids expensive ^{13}C-labeling and it gives access to orientational information for large proteins that strongly aggregate at concentrations above 50 μM, i.e., experimental conditions where 3D triple resonance experiments are not sensitive enough to allow backbone resonance assignment.

Introduction

The wealth of genomic data that has recently become available has created a need for rapid and efficient structural characterization of the corresponding proteins (Lander et al., 2001; Venter et al., 2001). As protein interactions play pivotal roles in a wide variety of biological processes, especially the determination of structures of protein-protein and protein-ligand complexes is desired. Structural characterization of protein complexes is traditionally difficult by both X-ray crystallography and NMR spectroscopy and only recently accurate structure determination of protein complexes became feasible using NMR derived residual dipolar couplings (Clore, 2000; Tjandra and Bax, 1997; Tjandra et al., 1997).

*E-mail: mzwecks@gwdg.de

Residual dipolar couplings (RDCs) can be observed in solution when a molecule is aligned with the magnetic field, either as a result of its own magnetic susceptibility anisotropy (BothnerBy, 1996; Gayathri et al., 1982), or caused by an anisotropic environment (Emsley, 1996; Saupe and Englert, 1963; Tjandra and Bax, 1997). If well-defined structures of either complete macromolecules or even smaller fragments are available, an alignment tensor that describes the average orientation of this structure with respect to the magnetic field can be calculated from RDCs (Losonczi et al., 1999; Tjandra et al., 1996). These molecular alignment tensors open up a wide range of possibilities: They can be used to determine the relative orientation of protein domains facilitating structure determination of weak affinity complexes in solution (Clore, 2000; Tjandra et al., 1997). They enable determination of the global helical structure of RNA

(Mollova et al., 2000; Vermeulen et al., 2000) and they can be used to study dynamic effects (Fischer et al., 1999; Meiler et al., 2001; Peti et al., 2002; Tolman, 2002; Tolman et al., 2001). Analysis of the relation between a molecule's 3D shape and its alignment tensor values can yield important insights into all of these processes (Zweckstetter and Bax, 2000).

It is important to stress that these new methods are based on high-resolution structures of free molecules or high-quality models as in the case of RNA duplexes with regular A-form geometry. Therefore, the dipolar couplings itself are not required any more to determine these structures, but just serve to extract the important orientational information encoded by a molecular alignment tensor. However, in order to determine an alignment tensor protein backbone resonances have to be assigned first. This assignment process is particularly difficult and time consuming for protein sizes above 20 kDa (Moseley and Montelione, 1999). Recently, it was shown that RDCs can also be used for resonance assignment. They are very effective for reducing chemical shift degeneracies (Zweckstetter and Bax, 2001), they allow simultaneous resonance assignment and structure determination in case of small proteins (Tian et al., 2001) and they can enhance assignment of proteins and RNA when a structure or high-quality structural model is available (Al-Hashimi et al., 2002; Hus et al., 2002).

Here we show that, when protein backbone resonances belonging to a specific amino acid type are selected either by amino acid specific labeling, by their characteristic C^{α}/C^{β} chemical shifts or by dedicated NMR experiments, molecular alignment tensors of large proteins, such as the 370 residue maltodextrin-binding protein, can be determined without prior backbone resonance assignment. This offers the opportunity to greatly accelerate determination of three-dimensional structures of protein-protein and protein-ligand complexes and validation of multimeric states of proteins.

Methods

For a pair of spin-1/2 nuclei, P and Q, separated by a distance r_{PQ}, the dipolar coupling D^{PQ} is related to the average orientation of the whole molecule by

$$D^{PQ} = -\mu_0\gamma_P\gamma_Q h/(8\pi^3 < r_{PQ}^3 >) \sum_{i,j} A_{ij} \cos\phi_i^{PQ} \cos\phi_j^{PQ}, \quad (1)$$

where **A** is a second-rank alignment tensor, γ_P and γ_Q are the gyromagnetic ratios, h is Planck's constant, μ_0 is the magnetic permeability of vacuum, r_{PQ} is the internuclear distance, and ϕ_i^{PQ} is the angle between the P–Q internuclear vector and the ith molecular axis. Using Equation (1) the alignment tensor **A** can conveniently be derived from experimental dipolar couplings and a known structure, either by singular value decomposition (SVD) (Losonczi et al., 1999) or by an iterative, multi-dimensional least-squares minimization of the difference between experimental and back-calculated RDCs (Tjandra et al., 1996). SVD is more stable than iterative least-squares minimization but does not permit certain parameters, such as the alignment tensor magnitude D_a or rhombicity R, to be held constant. Fits carried out in this study are based on SVD.

The uncertainty of alignment tensor values determined by exhaustive back-calculation is evaluated by addition of random noise to selected bond vector orientations, the so-called structural noise Monte-Carlo approach (Zweckstetter and Bax, 2002). The amplitude of structural noise is automatically adjusted by the program PALES to match the rmsd between experimental and back-calculated dipolar couplings and 1000 cycles are performed in the Monte-Carlo procedure. Exhaustive alignment tensors (without known assignment) are compared to alignment tensors obtained by SVD (with known assignment) using the generalized angle ϑ (Sass et al., 2001). In addition, the deviation of the principal axis of the exhaustive alignment tensor, the z axis, from its SVD orientation is considered. For visual inspection of tensor orientations Sauson-Flaumsteed projections are used (Bugayevskiy and Snyder, 1995).

Exhaustive back-calculation (exhSVD) has been implemented into the dipolar coupling analysis software PALES (M. Zweckstetter, unpublished) using the C programming language. The new module is accessible by calling PALES with the command-line argument '*-exhaust*'. Missing RDCs or peaks can be accommodated by addition of dummy RDCs (dipolar coupling value 888.00 Hz) to the dipolar coupling input table.

The approach is demonstrated on three proteins for which experimental dipolar couplings have been reported and a high-resolution crystal structure is available: ubiquitin (76 aa; PDB codes: 1UBQ [1.8 Å] and 1AAR [2.3 Å]; NMR restraints: 1D3ZMR) (Cornilescu et al., 1998; Vijaykumar et al., 1987), two-domain maltodextrin-binding protein (MBP) (370 aa;

Table 1. Summary of SVD-fit of experimental RDCs to known crystal structures for four proteins used in this study

PDB	N_{aa}[a]	Res. (Å)[b]	N_{dc}[c]	r_{NH}[d]	r_{NC}[e]	$r_{C'CA}$/ $r_{HNC'}$[f]	D_a (Hz)[g]	D_a^{Min}/ D_a^{Max} (Hz)[h]	R[i]	R^{Min} / R^{Max} [j]	$\Delta\vartheta$ (°)[k]	$\Delta\phi_z$ (°)[l]
1UBQ	76	1.8	69	0.97	0.98	0.95	−9.1	−9.8/ −8.4	0.15	0.08/0.21	8	5
1AAR	76	2.3	69	0.95	0.97	0.94	−8.8	−9.8/−8.0	0.11	0.05/0.27	11	6
1CI4	89	1.9	65	0.97	0.92	0.87	−14.0	−15.3/−12.9	0.16	0.08/0.35	10	5
1OMP	370	1.8	279	0.93	0.96	0.94	−15.8	−16.8/−14.6	0.16	0.10/0.25	6	3
1D8C$^+$	723	2.0	656	0.95	0.95	0.93	−21.5	−22.2/−20.8	0.42	0.38/0.46	4	2

[a] N_{aa}: # of residues.
[b] Resolution of crystal structure.
[c] N_{dc}: # of experimental $^1D_{NH}$ couplings.
[d] r_{NH}: Pearson's linear correlation coefficient between experimental and back-calculated $^1D_{NH}$ RDCs.
[e] r_{NC}: Correlation between experimental and back-calculated $^1D_{NC'}$ RDCs.
[f] $r_{C'CA}/r_{HNC'}$: Correlation between experimental and back-calculated $^1D_{C'CA}$ (or $^2D_{HNC'}$ for 1UBQ) RDCs.
[g] Magnitude of alignment tensor normalized to $^1D_{NH}$.
[h] Error bounds for D_a: Maximum and minimum value as obtained from the 'structural noise Monte-Carlo method'.
[i] Rhombicity of alignment tensor.
[j] Error bounds for R: Maximum and minimum value as obtained from the 'structural noise Monte-Carlo method'.
[k] Uncertainty in the general orientation ϑ of the alignment tensor as evaluated by the 'structural noise Monte-Carlo method'.
[l] Uncertainty in the orientation of the z-axis of the alignment tensor as evaluated by the 'structural noise Monte-Carlo method'.
$^+$ RDCs for MSG were predicted from the shape of the crystal structure 1D8C. The effect of structural uncertainty was simulated as described in the text.

PDB code: 1OMP [1.8 Å]; NMR restraints: 1EZPMR) (Mueller et al., 2000; Sharff et al., 1992) and homodimeric barrier-to-autointegration factor (BAF) (89 aa; PDB code: 1CI4 [1.9 Å]; NMR restraints: 2EZXMR) (Cai et al., 1998; Umland et al., 2000). Protons were added to crystal structures using Molmol (Koradi et al., 1996). For all three proteins dipolar couplings were used as deposited in the Protein Data Bank (PDB). As can be seen from Table 1 experimental RDCs fit well to published crystal structures. In addition, exhSVD was tested on the 723 residue protein malate synthase G (PDB code: 1D8C [2.0 Å]) (Howard et al., 2000). No experimental RDCs have so far been reported for MSG. Therefore, $^1D_{NH}$, $^1D_{NC'}$ and $^1D_{C'CA}$ RDCs were calculated from the crystal structure based on the alignment tensor that is predicted from the shape of MSG using the program PALES. The effect of structural uncertainty was taken into account by slightly reorienting the selected internuclear vectors according to a Gaussian cone-shaped distribution with a standard deviation of 10° (Zweckstetter and Bax, 2002).

Results and discussion

Particularly useful for determination of a molecular alignment tensor best-fit to experimental RDCs is singular value decomposition (SVD), as it is very fast and requires a minimum of only five RDCs (Losonczi et al., 1999). In our approach we build on these favorable properties of SVD and on the fact that amino acid types of NMR resonances can efficiently be identified by a variety of methods. In the first step all NMR resonances belonging to a specific amino acid type are identified. This identification is achieved preferably by amino acid specific labeling using auxotrophic strains. The identification using auxotrophic strains is accurate (i.e., in most cases all expected HN-N correlations are observed and no cross labeling is present). It requires only ^{15}N labeled protein and can be performed even without purification of the protein, i.e., with crude cell lysates from as little as 100 ml cultures, thus saving time and money (Gronenborn and Clore, 1996; Ou et al., 2001). Alternatively, the C^α chemical shift allows to uniquely identify glycine residues, and when the C^β chemical shift is also available alanine, threonine and serine residues can also be identified unambiguously (Grzesiek and Bax, 1993; Spera and Bax, 1991). In addition, a variety of NMR experiments have been devised that allow selection of specific amino acid types (Dotsch et al., 1996; Schubert et al., 2001).

After identification of residues of a specific amino acid type (for example the 8 tryptophans in MBP; $n_{total} = 8$) an exhaustive search for all possible assignments is performed ($n_{total}!$ steps) and for each assignment a SVD calculation is done ('exhaustive back-calculation'; exhSVD). For each of these SVD

Table 2. Comparison of alignment tensors determined by exhSVD with those obtained by assignment-based SVD for two different crystal structures of ubiquitin

	N_{aa}[a]	N_{pk}[b]	N_{dc}[c]	N_{wr}[d]	D_a (Hz)[e,⊥]	R[f]	ϑ (°)[g]	$\Delta\vartheta$ (°)[h]	D_a^{ex} (Hz)[+]	R^{ex}	ϑ^{ex} (°)	$\Delta\vartheta^{ex}$ (°)	ϑ^{svd-ex} (°)[i]
							1UBQ (1.8 Å)						
ALA	2	2	6	0	−10.0	0.11	8	76	−10.0	0.11	8	76	0
ARG	4	4	12	2	−6.7	0.14	37	75	−8.2	0.12	70	26	35
ASN	2	2	6	0	18.5	0.08	52	141	18.5	0.08	52	141	0
ASP	5	5	15	0	−9.7	0.20	8	22	−9.7	0.20	8	22	0
GLN	6	6	18	0	−9.7	0.17	8	14	−9.7	0.17	8	14	0
GLU	6	5	15	0	−8.9	0.23	5	18	−8.9	0.23	5	18	0
GLY	6	5	12	0	−7.5	0.11	15	29	−7.5	0.11	15	29	0
ILE	7	7	21	2	−9.0	0.15	5	17	−9.1	0.18	4	15	2
LEU	9	9	27	2	−9.4	0.17	6	17	−9.4	0.17	6	17	2
LYS	7	7	21	0	−9.3	0.03	15	13	−9.3	0.03	15	13	0
SER	3	3	9	0	−17.6	0.47	28	123	−17.6	0.47	28	123	0
THR	7	6	16	0	−8.6	0.25	8	17	−8.6	0.25	8	17	0
VAL	4	4	12	0	−8.8	0.24	14	13	−8.8	0.24	14	13	0
							1AAR (2.3 Å)						
ALA	2	2	6	0	−10.2	0.14	7	129	−10.2	0.14	7	129	0
ARG	4	4	12	2	−6.4	0.19	28	68	−7.5	0.16	59	26	32
ASN	2	2	6	0	21.6	0.05	58	120	21.6	0.05	58	120	0
ASP	5	5	15	0	−10.7	0.20	12	28	−10.7	0.20	12	28	0
GLN	6	6	18	0	−8.5	0.03	7	24	−8.5	0.03	7	24	0
GLU	6	5	15	0	−8.5	0.13	6	13	−8.5	0.13	6	13	0
GLY	6	5	12	3	−11.2	0.42	17	106	−12.4	0.43	23	40	11
ILE	7	7	21	4	−10.3	0.16	16	17	−10.2	0.17	13	14	3
LEU	9	9	27	4	−9.4	0.14	12	17	−9.9	0.06	20	16	10
LYS	7	7	21	2	−7.6	0.12	7	31	−7.5	0.24	12	31	9
SER	3	3	9	0	−10.9	0.21	15	101	−10.9	0.21	15	101	0
THR	7	6	16	4	−9.2	0.17	4	17	−9.0	0.23	10	11	7
VAL	4	4	12	0	−8.3	0.20	13	29	−8.3	0.20	13	29	0

[⊥] Alignment tensor values in columns 6 to 9 are obtained by assignment-based SVD.
[+] Alignment tensor values in columns 10 to 13 are obtained by exhaustive back-calculation ('ex').
[a] N_{aa}: # of residues with the specific amino acid type present in ubiquitin.
[b] N_{pk}: # of residues with the specific amino acid type for which dipolar couplings were reported.
[c] N_{dc}: Total # of RDCs available for the specific amino acid type.
[d] N_{wr}: # of wrong assignments present for the permutation with the overall lowest rmsd.
[e] Magnitude of alignment tensor normalized to $^1D_{NH}$.
[f] Rhombicity of alignment tensor.
[g] Deviation of alignment tensors calculated for residues belonging to a specific amino acid type from the alignment tensor calculated using all 208 RDCs available for ubiquitin.
[h] Spread in ϑ is evaluated by the 'Monte-Carlo structural noise' method. Note that $\Delta\vartheta$ indicates the estimated uncertainty, whereas ϑ is the true error.
[i] Deviation of alignment tensors determined by exhSVD from those obtained by SVD; only RDCs of a specific amino acid type are employed in both cases.

calculations the rmsd between experimental and back-calculated RDCs is determined. These rmsd values are then used to rank all SVD calculations and the alignment tensor with the smallest rmsd is selected. Due to the high speed of SVD calculations, an exhaustive search can be performed for a set of up to 10 amino acids ($10! = 3.6288e+06$ assignment possibilities) in less than five minutes on a single processor SGI origin (270 MHz) workstation. More residues of a given type can be used if supercomputers are accessible.

The limit of about 10 residues of a given type is due to the explosion of assignment possibilities with increasing number of residues (n!). However, with a slightly modified approach this limit can be lifted by about five residues. Assume we measured $^1D_{NH}$ RDCs for the twelve serines ($n_{total} = 12$) in MBP. From these twelve $^1D_{NH}$ RDCs we select a subset of six ($n_{sel} = 6$) couplings. In order to perform a SVD calculation we need the orientation of six HN-N vectors. However, we have a total of twelve HN-N vectors in the crystal structure of MBP and we do not know which six vectors to choose. Therefore, two exhaustive searches have to be performed. For a specific selection of six HN-N vectors all possible assignments ($n_{sel}!$ steps) to the six $^1D_{NH}$ RDCs are tested for minimum rmsd. In addition, a second exhaustive search is performed for all possibilities to pick six from the total of twelve HN-N vectors. This results in a total number of $n_{total}! / (n_{total} - n_{sel})!$ SVD calculations. When a subset of six couplings ($n_{sel} = 6$) is selected, amino acid types with up to 15 residues ($n_{total} = 15$) are therefore accessible by exhSVD. Note, however, that RDCs for only six residues can be used decreasing the robustness of exhSVD (see below for a detailed discussion of the robustness of exhSVD).

The performance of exhSVD is evaluated for ubiquitin using both the 1.8 Å and the 2.3 Å crystal structure. Table 2 compares alignment tensors obtained by exhSVD with those determined by SVD. Whereas for a regular SVD calculation the assignment has to be known, in exhSVD only information about the amino acid type is required. So, for example, if the assignment of ubiquitin would not be known, its six glycine residues could be identified by amino acid specific labeling or by their characteristic $^{13}C_\alpha$ chemical shift. The RDCs measured for these six glycines can then be used for exhSVD. In order to evaluate, how alignment tensor values determined by exhSVD are influenced by the specific amino acid type, Table 2 shows the results of exhSVD for all amino acid types for which more than a total of five $^1D_{NH}$, $^1D_{NC'}$ and $^2D_{HNC'}$ RDCs could be measured. Most abundant in ubiquitin are leucines ($n_{total} = 9$) and 27 RDCs are available. Using the 1.8 Å crystal structure the alignment tensor obtained by exhSVD from these 27 RDCs is highly accurate. Its orientation makes a generalized angle of just 6° with the alignment tensor determined by SVD. This corresponds to just 2° in the orientation of the z axis. The magnitude $D_a(NH) = -9.4$ Hz and the rhombicity $R = 0.17$ determined by exhSVD are also in agreement with SVD values (see Table 1). For other amino acid types with more than 15 RDCs, such as ASP, GLN or THR, similar results are obtained and the generalized angle with respect to the SVD orientation is below 15°. This corresponds to an error in the orientation of the principal axis of the alignment tensor, the z-axis, of less than 8° demonstrating that it is possible to determine the molecular alignment tensor of ubiquitin without prior backbone assignment.

Errors in alignment tensor values, however, increase when only a small number of RDCs is available.

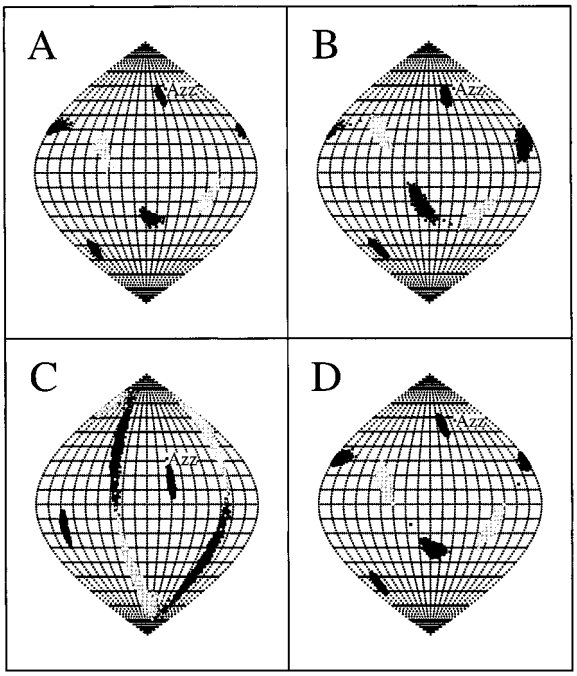

Figure 1. Orientation of the molecular alignment tensor of ubiquitin determined without knowledge of its backbone assignment using exhaustive back-calculation on ILE (A), LEU (B), ARG (C) and ILE-ARG (D). Experimental $^1D_{NH}$, $^1D_{NC'}$, $^2D_{HNC'}$ couplings observed in 30:10 DMPC:DHPC bicelles and a 1.8 Å crystal structure (PDB code: 1UBQ) are used (Table 2). Orientations are visualized using Sauson-Flaumsteed projections (z-axis/black with label Azz; y-axis/black; x-axis/gray). Uncertainty estimates are obtained from 'structural noise Monte-Carlo simulations'.

For SER, for example, only a total of nine RDCs are available and the alignment tensor already differs by $\vartheta = 28°$ from the SVD orientation. The situation is even worse for ARG and ASN where the generalized angle with respect to the correct orientation amounts to ~60°. This is not any more a slight deviation from the true orientation, but rather has to be regarded as an incorrect alignment tensor orientation (see also Figure 1). In addition, alignment tensor uncertainties are prone to increase when the quality of the reference structure decreases. For example, the error in alignment tensor orientation obtained by exhSVD on LEU increases from 6 to about 20° when the 2.3 Å instead of the 1.8 Å crystal structure is used (Table 2). Therefore, it is essential to evaluate the uncertainty in alignment tensor values obtained by exhSVD.

Evaluation of uncertainty

Evaluation of uncertainties is not only important for exhSVD when the assignment is not known, but also in case of SVD with known assignment. Even in regular SVD calculations the uncertainty in alignment tensor values strongly increases when going to less than 15 RDCs or when the quality of the structure decreases (Zweckstetter and Bax, 2002). In order to evaluate these uncertainties we recently introduced a so-called 'structural noise Monte-Carlo method' as implemented in PALES (Zweckstetter and Bax, 2002). This method is based on the observation that errors in alignment tensor values obtained from one- and two-bond RDCs are dominated by uncertainties in the reference structure. It evaluates these errors by addition of random noise to selected bond vector orientations. To make it applicable for exhSVD, we assume that the assignment from the exhaustive search with the smallest rmsd corresponds to the true backbone assignment. This assumption is clearly not correct as exchange of two residues with equal RDCs does not affect agreement between an assumed geometry and RDCs and can therefore lead to partially wrong assignments (Table 2; column N_{wr}). For example, RDCs measured for LEU 15 ($^1D_{NH} = 11.0$ Hz, $^1D_{NC'} = 0.3$ Hz, $^2D_{HNC'} = 2.3$ Hz) and LEU 56 ($^1D_{NH} = 9.6$ Hz, $^1D_{NC'} = 0.5$ Hz, $^2D_{HNC'} = 1.9$ Hz) are highly similar. In combination with uncertainties in the 1.8 Å crystal structure of ubiquitin this results in a lower rmsd for the assignment where LEU 15 and LEU 56 are exchanged compared to the correct assignment. The alignment tensor, however, is almost unaffected by the switched RDCs with a change in orientation of $\vartheta < 2°$. Assuming that the assignment from the exhaustive search with the smallest rmsd corresponds to the true backbone assignment, we perform the 'structural noise Monte-Carlo method'. Variations in alignment tensor values during the Monte-Carlo simulation are then used as estimates for their uncertainty. Table 2 lists these uncertainty estimates for the exhSVD results together with the actual deviation from SVD values. For ASN and LEU, for example, the deviation of the exhSVD alignment tensor from its true orientation is $\vartheta = 52°$ and $\vartheta = 6°$, respectively, and the corresponding Monte-Carlo spreads are 141° and 17°. This demonstrates that spreads obtained from 'structural noise Monte-Carlo simulations' can be used to evaluate uncertainties in exhSVD alignment tensors.

A problematic case is ARG. There are a total of four ARG in ubiquitin and two of them (ARG 72 and 74) are in the highly flexible C-terminus with NMR N-H order parameters $S^2 < 0.6$ (Tjandra et al., 1995). For these two residues the assumed geometry that is obtained from the crystal structure is incorrect. This makes back-calculation prone to errors as SVD is based on the assumption of a known geometry. Accordingly, the rmsd between experimental RDCs and those obtained by regular SVD (with known assignment) is high and the uncertainty in alignment tensor orientation amounts to 75° despite the fact that a total of 12 RDCs is available (Table 2). Exhaustive back-calculation is even more difficult in such a case. In exhSVD all possible assignments are evaluated for best-fit to an assumed geometry. However, due to errors in the reference structure the rmsd between experimental RDCs and those calculated from the true alignment tensor is very high. Therefore, RDCs calculated from an incorrect alignment tensor can potentially fit better to experimental couplings especially when an exhaustive search for all possible assignments is performed. This is the case for exhSVD using the four ARG of ubiquitin. Due to errors in the assumed geometry of the C-terminus, an alignment tensor that deviates by $\vartheta = 70°$ from the true orientation can accurately reproduce experimental RDCs. In this best-fit the assignment of ARG 42 is exchanged with ARG 72 and the residual rmsd points to an accurate alignment tensor. Based on this residual rmsd the spread obtained from the 'structural noise Monte-Carlo method' is much too small (26°) and significantly underestimates the true uncertainty. Thus, when systematic differences between the structure in solution and in the crystalline state are present, Monte-Carlo spreads no longer reliably indicate uncertainties in alignment

tensor values. This is true for both regular SVD and exhSVD as the 'structural noise Monte-Carlo method' only takes into account random structural noise. For exhSVD, however, systematic structural errors are especially problematic due to the different assignment possibilities. In this respect, it is also important to note that alignment tensor orientations cannot be extracted by exhSVD, if the distribution of internuclear vectors is completely isotropic, as all assignment permutations are then equivalent. Therefore, the decision which amino acid to label selectively and to use for exhSVD should be based on the known crystal structure: selected internuclear vectors should be sufficiently anisotropic and most of the selected residues should be part of regular secondary structure. This reduces the likelihood of structural rearrangements and minimizes the influence of dynamical averaging of RDCs in solution. In addition, a maximum number of RDCs is always recommended. This increases the stability of exhSVD and results, as for SVD (Zweckstetter and Bax, 2002), in more accurate alignment tensors.

exhSVD alignment tensors can further be validated if two different amino acid types can be identified by amino acid specific labeling, by their characteristic C^{α}/C^{β} chemical shifts or by dedicated NMR experiments. In this validation separate exhSVD calculations are performed for the two different amino acid types. If the results from both calculations agree within their respective Monte-Carlo spreads, the alignment tensor is validated. This is demonstrated in Figure 1 for ubiquitin. Figures 1A to 1B show the orientation of alignment tensors determined by exhSVD for ILE and LEU, respectively. The alignment tensor orientations for the two amino acid types agree within their estimated uncertainties making this alignment tensor orientation highly reliable. For ARG, on the other hand, the alignment tensor orientation determined by exhSVD (Figure 1C) deviates significantly from that observed in Figures 1A and 1B. This indicates that exhSVD results on ARG are influenced by structural rearrangements in solution and cannot be trusted (as could have been anticipated from the high B-factors of residue 72-76 in the crystal structure). After the alignment tensor has been validated RDCs for different amino acid types can be used simultaneously in exhSVD. In this exhaustive search permutations are restricted to the correct amino acid type. So, if exhSVD is performed using both ILE and ARG, RDCs measured on ILE can only be assigned to ILE internuclear vectors and the same for ARG couplings. Combining two amino acid types has the advantage that the number of assignment possibilities is reduced and more RDCs are available for exhSVD in total. This makes exhSVD more robust and decreases uncertainties in alignment tensor values. This is shown in Figure 1D for the combination of ARG with ILE. The incorrect geometry of the C-terminus of ubiquitin does no longer cause a wrong alignment tensor, as no ILE residues are located in the C-terminus and both ILE and ARG dipolar couplings have to be best-fit to a single alignment tensor.

A final validation can be performed if $^{1}D_{NH}$ RDCs on the whole protein are measured using simple 2D HSQC or TROSY sequences. From the distribution of these RDCs the magnitude and rhombicity of the alignment tensor can be obtained (Clore et al., 1998). These values have to agree with those obtained from exhaustive back-calculation. In this way, for example, the alignment tensor values of SER and ASN are identified as unreliable (Table 2).

Increasing alignment tensor accuracy

Applications based on molecular alignment tensors depend on the accuracy of these tensors. Therefore, it is desired to determine the alignment tensor values as accurate as possible. As for SVD, the accuracy of alignment tensor values depends critically on the quality of the reference structure. Accordingly, the selection of an amino acid type that is used for exhaustive back-calculation should be based on the available three-dimensional structure (see above). In addition, as many RDCs as possible should be used. For example, when exhSVD is performed on the six glycines in ubiquitin and only $^{1}D_{NH}$ and $^{1}D_{NC'}$ instead of $^{1}D_{NH}$, $^{1}D_{NC'}$ and $^{2}D_{HNC'}$ couplings are included, the alignment tensor deviates by a generalized angle of 72° from its correct orientation (Table 3). Similar to ARG discussed above this is due to the wrong geometry of the C-terminus where two of the six glycines are located (GLY 75 and GLY 76). Inclusion of $^{2}D_{HNC'}$ (or $^{1}D_{C'CA}$) couplings overcomes this problem and with a deviation of $\vartheta = 15°$ from the true orientation the alignment tensor is accurately determined by exhSVD (Table 3; row 2).

Instead of using additional types of RDCs, dipolar couplings can be measured in a second alignment medium. In this case two SVD calculations are done for each permutation during the exhaustive search, i.e. one for each alignment medium. Permutations are then ranked according to the sum of the rmsds ob-

Table 3. Effect of an increasing number of experimental constraints on the robustness of exhSVD[⊥]

	N_{aa}	N_{pk}	N_{dc}	N_{wr}	D_a (Hz)	R	ϑ (°)	$\Delta\vartheta$ (°)	D_a^{ex} (Hz)	R^{ex}	ϑ^{ex} (°)	$\Delta\vartheta^{ex}$ (°)	$\vartheta^{svd\text{-}ex}$ (°)
						1UBQ (1.8 Å)							
GLY●	6	5	9	4	−7.6	0.08	16	35	16.4	0.14	72	13	69
GLY+	6	5	12	0	−7.5	0.11	15	29	−7.5	0.11	15	29	0
GLY°	6	5	18	0	−7.6	0.08	16	35	−7.6	0.08	16	35	0
GLY-ASPx	11	10	27	0	−8.9	0.14	7	16	−8.9	0.14	7	16	0

[⊥]For a description of parameters see Table 2.
●$^1D_{NH}$, $^1D_{NC'}$ observed in 30:10 DMPC:DHPC bicelles.
+$^1D_{NH}$, $^1D_{NC'}$ and $^2D_{HNC'}$ couplings.
°$^1D_{NH}$, $^1D_{NC'}$ in two alignment media (30:10 DMPC:DHPC and 30:10:1 DMPC:DHPC:CTAB). RDCs in [30:10:1 DMPC:DHPC:CTAB] have been rescaled, such that the magnitude of alignment is the same in the two media.
x$^1D_{NH}$, $^1D_{NC'}$ and $^2D_{HNC'}$ couplings from GLY and ASP are used simultaneously in exhSVD.

Table 4. Comparison of alignment tensors determined by assignment-free exhSVD with those obtained by assignment-based SVD for the 370 residue maltodextrin-binding protein[⊥]

	N_{aa}	N_{pk}	N_{dc}	N_{wr}	D_a (Hz)	R	ϑ (°)	$\Delta\vartheta$ (°)	D_a^{ex} (Hz)	R^{ex}	ϑ^{ex} (°)	$\Delta\vartheta^{ex}$ (°)	$\vartheta^{svd\text{-}ex}$ (°)
ARG	6	5	15	1	−17.2	0.34	12	34	−19.6	0.36	14	27	6
GLN	9	8	24	2	−17.0	0.20	8	20	−17.0	0.20	8	20	1
HIS	3	2	6	1	−16.0	0.13	37	113	−22.9	0.34	15	26	45
MET	6	5	13	0	−16.1	0.12	13	15	−16.1	0.12	13	15	0
TRP	8	6	17	0	−15.3	0.21	3	29	−15.3	0.21	3	29	0

[⊥]For a description of parameters see Table 2.

tained from these two SVD calculations. For ubiquitin, for example, RDCs were measured in nearly neutral (30:10 DMPC:DHPC) as well as in slightly positively charged bicelles (30:10:1 DMPC:DHPC:CTAB) (Cornilescu et al., 1998). Addition of positive or negative charges to bicelles results in an electrostatic interaction between ubiquitin and the bicelle particles and leads to changes in observed RDCs (Ramirez and Bax, 1998). This reduces degeneracy in experimental RDCs and favors the permutation in the exhaustive search that corresponds to the correct assignment. For GLY the best-fit between experimental $^1D_{NH}$ and $^1D_{NC'}$ couplings (observed in nearly neutral bicelles) and those back-calculated from the structure is obtained when four of the six glycines are not assigned to their correct internuclear vectors. On the other hand, when RDCs from the second alignment medium are included, the lowest overall rmsd is observed for the correct assignment and the alignment tensor determined by exhSVD is the same as that determined by SVD (Table 3; row 3).

Especially useful for exhaustive back-calculation is the usage of two or more different amino acid types. As explained above, separate exhSVD calculations on two different amino acid types allow validation of alignment tensor values and make exhSVD calculations more robust. In addition, more RDCs are available for back-calculation in total. This increases the accuracy of alignment tensor values as is shown by the combined exhSVD on GLY and ASP. Using $^1D_{NH}$, $^1D_{NC'}$ and $^2D_{HNC'}$ couplings identical results are obtained for exhSVD and for SVD. The alignment tensor deviates by $\vartheta = 8°$ and $\vartheta = 15°$ from the correct alignment tensor orientation for ASP and GLY, respectively. The corresponding uncertainties are 22° and 29°. Combination of RDCs of GLY and ASP in the exhaustive back-calculation results in a total of 27 RDCs and decreases the deviation to $\vartheta = 7°$ (4° for the z-axis) and the uncertainty estimate to 16° (8° for the z-axis) (Table 3; row 4).

Assignment-free determination of alignment tensors for proteins above 40 kDa

exhSVD determines the molecular alignment tensor of a protein by performing an exhaustive search for a subset of residues that were identified by their specific amino acid type. It can be applied as long as at

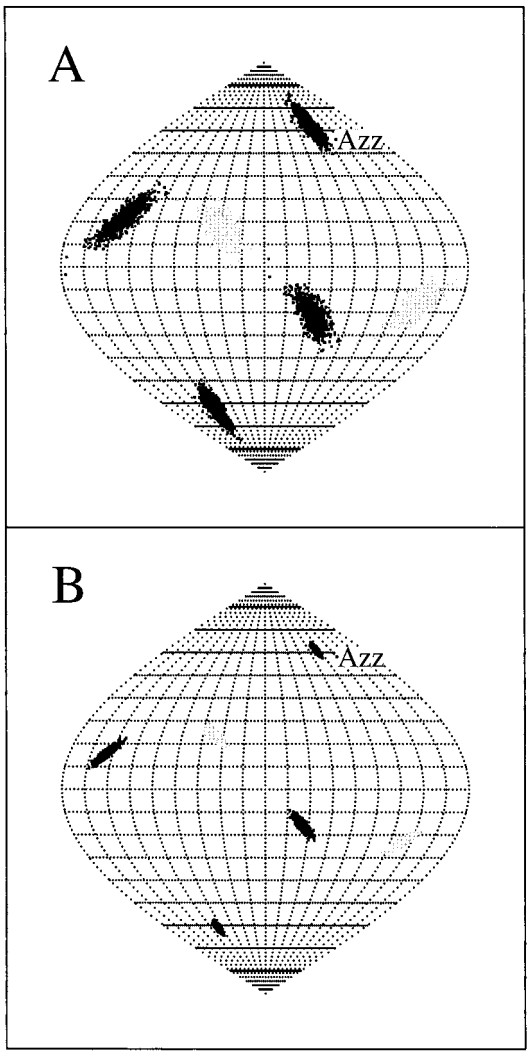

Figure 2. Orientation of the molecular alignment tensor of the 370 residue maltodextrin-binding protein (MBP) determined without (A) and with (B) knowledge of its backbone assignment. (A) Exhaustive back-calculation using 17 experimental $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C'CA}$ couplings of six observable tryptophans in MBP (Table 4); (B) regular SVD with 279 $^1D_{NH}$ couplings. Orientations are visualized using Sauson-Flaumsteed projections (z-axis/black with label Azz; y-axis/black; x-axis/gray). Uncertainty estimates are obtained from 'structural noise Monte-Carlo simulations'.

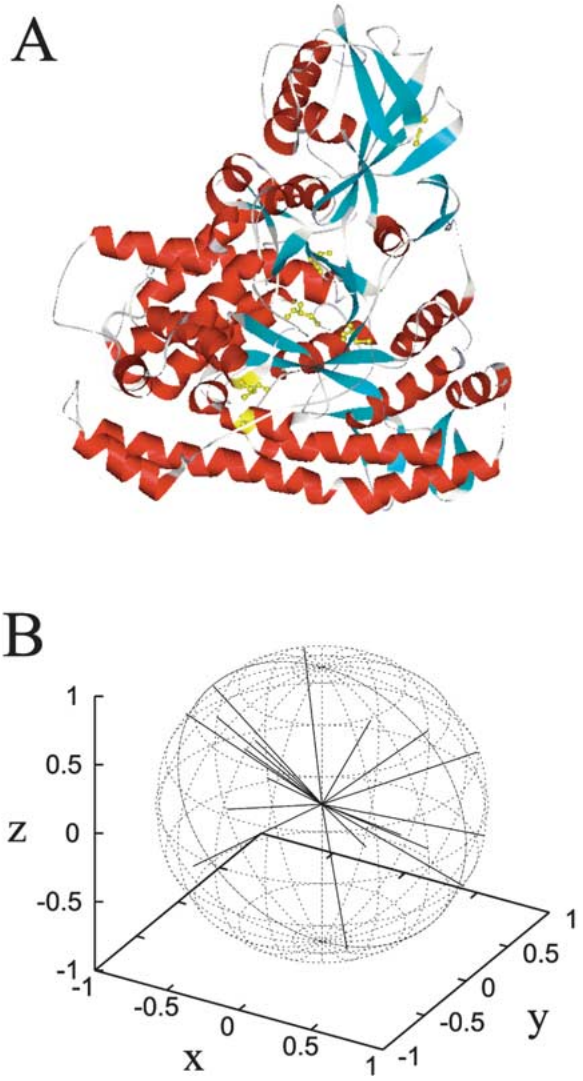

Figure 3. (A) Three-dimensional structure of the 723 residue protein malate synthase G (PDB code: 1D8C). Malate synthase G contains only six cysteines. These are highlighted in yellow including their side-chains. (B) Orientation of the 18 internuclear vectors in malate synthase G that correspond to the 18 $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C'CA}$ couplings of the six cysteines. Orientations are shown in the coordinate frame of the crystal structure (A).

least one amino acid type is less than about 11 times present in the primary sequence of the protein. Therefore, the actual size of the protein is not important: the approach works in the same way independent of whether the protein has a total size of 80 residues or 800 residues provided that the quality of the reference structure is the same. This is demonstrated for the 370 residue maltodextrin-binding protein (MBP) and the 723 residue protein malate synthase G (MSG).

MBP is a 370 residue protein consisting of two domains (Sharff et al., 1992; Skrynnikov et al., 2000). The assignment of MBP in 1998 significantly extended the size of proteins accessible to backbone resonance assignment and MBP is still one of the largest single-chain proteins assigned so far (Gardner et al., 1998). With respect to exhaustive back-calculation,

however, MBP is not more difficult than ubiquitin. In MBP five amino acid types, ARG, GLN, HIS, MET and TRP are less than 10 times present and each of them can readily be used for exhSVD. In addition, a 1.8 Å crystal structure is available (Sharff et al., 1992) and $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C'CA}$ can be measured accurately (Yang et al., 1999). Figure 2 compares the alignment tensor orientations of MBP obtained by regular SVD (based on the published assignment) and by exhaustive back-calculation on TRP. It shows that exhSVD is able to determine the alignment tensor of MBP without the need for backbone resonance assignment. Similar results are obtained when ARG, GLN, HIS or MET are used instead of TRP with an average generalized angle of $\sim 10°$ between the exhSVD and the SVD tensor and an uncertainty of $\sim 23°$ (Table 4). This corresponds to an error of less than 7° in the orientation of the principal axis of the alignment tensor. The bigger uncertainties in exhSVD values compared to those of SVD on the whole molecule are due to the lower number of RDCs used (17 RDCs for exhSVD versus 279 $^1D_{NH}$ for SVD). Application of exhSVD to MBP is especially challenging, as the relative orientation of its two domains in solution is changed by $\sim 10°$ compared to its crystal structure (Skrynnikov et al., 2000). Many bigger proteins are composed of smaller domains that have common folds and for which large structural difference in solution and in the crystal state are not expected. Changes in the relative orientation of these domains caused by crystal packing, however, are more likely. Figure 2 and Table 4 show that such modest changes in domain orientation can be tolerated by exhSVD. Detection of such modest changes in relative domain orientation from a small number of residues as used in exhSVD, on the other hand, is difficult due to the limited accuracy of corresponding alignment tensors ($\sim 7°$ for the principal tensor axis in case of MBP).

Malate synthase G (MSG) is a 723 residue protein consisting of four domains. A 2.0 Å crystal structure of MSG is available (Howard et al., 2000). Recently, the sequential assignment of MSG was accomplished using TROSY 4D NMR spectroscopy (Tugarinov et al., 2002). This effectively doubled the size of proteins (number of residues in a single chain) that can be assigned by NMR. In order, however, to remove signal overlap, reduce chemical shift degeneracies and allow the assignment of a protein of that size, many three- and four-dimensional had to be recorded. This also included dedicated NMR experiments in order to unambiguously identify residues belonging to a specific amino acid type, similar to what would be needed for exhSVD. In total, 1.5 months of measurement time were required and the assignment process took about one month (Tugarinov et al., 2002). On the other hand, despite the total length of 723 residues, MSG contains only six cysteines (Figure 3) making it accessible to exhSVD. The main size limitation for exhSVD is the ability to accurately measure dipolar couplings. For proteins with a molecular weight of more than 60 kDa the upfield component in a $\{^1H\}$-^{15}N doublet relaxes very fast and measurement of RDCs is difficult (Kontaxis et al., 2000). However, with a lot of effort put into the development of new methods for measurement of RDCs this size limit can hopefully be lifted. In order to demonstrate the principal applicability of exhSVD for very big proteins, we here use simulated $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C'CA}$ couplings. These RDCs were calculated from the crystal structure using the alignment tensor predicted from the shape of MSG. This simulates the situation when RDCs are measured in MSG dissolved in a nearly neutral bicelle medium (Zweckstetter and Bax, 2000). In order to account for structural uncertainties, we slightly reoriented the selected internuclear vectors according to a Gaussian cone-shaped distribution with a standard deviation of 10°. A regular SVD calculation on these simulated RDCs and the noise-distorted structure results in a RDC correlation factor of 0.94 (Table 1). This resembles the situation that is commonly encountered for a 2.0 Å crystal structure and one-bond RDCs (see Table 1). Using these RDCs and the corresponding structure the alignment tensor of MSG is accurately predicted by exhaustive back-calculation. Its orientation deviates by $\vartheta = 8°$ from that determined by SVD (Figure 4). The error in the orientation of the z-axis is only 3°. The exhaustive search took less than one minute. This demonstrates that the accuracy of alignment tensors determined by exhSVD does not depend on protein size. Critical is the quality of the reference structure and the accuracy with which dipolar couplings can be measured.

Assignment-free alignment tensors from ^{15}N-labeled protein

In order to determine the backbone assignment of even small proteins in an automatic or semi-automatic manner the protein has to be labeled both with ^{15}N and with ^{13}C (Moseley and Montelione, 1999). Labeling with ^{13}C, however, is expensive and increases the costs of NMR in structural genomics. exhSVD, on the other

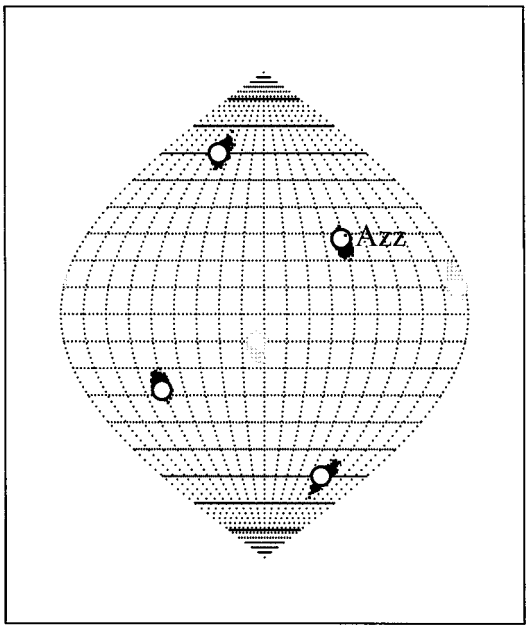

Figure 4. Orientation of the molecular alignment tensor of the 723 residue protein malate synthase G determined by exhaustive back-calculation on 18 $^{1}D_{NH}$, $^{1}D_{NC'}$, $^{1}D_{C'CA}$ couplings of six cysteines in MSG. The alignment tensor was determined without backbone resonance assignment. Orientations are visualized using Sauson-Flaumsteed projections (z-axis/black with label Azz; y-axis/black; x-axis/gray). Uncertainty estimates are obtained from 'structural noise Monte-Carlo simulations'. ○ Indicate the true orientation of the alignment tensor axes.

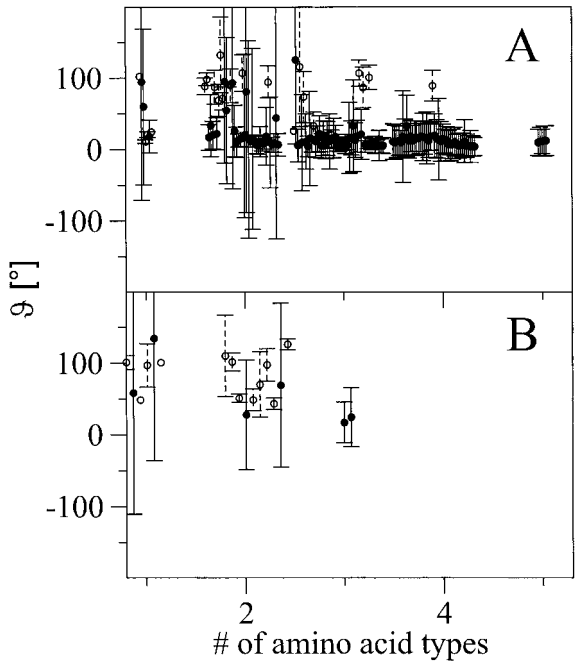

Figure 5. Deviation of alignment tensors determined by exhSVD from the correct orientation for an increasing number of selected amino acid types. Only $^{1}D_{NH}$ couplings from a single alignment medium are used. (A) exhSVD results on ubiquitin. (B) exhSVD results on maltodextrin-binding protein. Differences in tensor orientations are expressed by the generalized angle ϑ (Sass et al., 2001). Uncertainty estimates are obtained from 'structural noise Monte-Carlo simulations'. ● and ○ indicate exhSVD results where the error in the tensor orientation is smaller or larger than the estimated uncertainty, respectively. Note that in (A) (ubiquitin) only the worst case data are shown, i.e. where at least one amino acid type is strongly influenced by errors in the crystal structure.

hand, determines molecular alignment tensors without prior backbone assignment. This reduces the required measurement time, it avoids the assignment process and can be done even for large proteins such as MBP in a fully automatic manner. Due to errors in the crystal structure, however, three different types of dipolar couplings are usually required necessitating labeling with ^{13}C.

As demonstrated above exhSVD can be made more robust when residues of two different amino acid types are identified and used simultaneously in exhSVD. Figure 5 evaluates how the accuracy of alignment tensors increases when only ^{15}N-labeling is available and two, three or four amino acid types are identified by amino acid specific labeling. In order to further increase the robustness of exhSVD, we incorporate information about the magnitude and rhombicity of the alignment tensor into the exhaustive search. The magnitude and rhombicity of an alignment tensor can readily be determined from the extrema or the shape of a RDC histogram (Clore et al., 1998; Warren and Moore, 2001). During the exhaustive search we then allow only permutations for which SVD back-calculates an alignment tensor with a magnitude and rhombicity that is in agreement with the experimentally observed values. To evaluate agreement we use rather large error bounds, i.e., two times the uncertainty in D_a and R that is obtained from the histogram as part of a three-parameter non-linear least-square fit. This distinguishes between almost axially symmetric ($R < 0.3$) and highly rhombic alignment tensors ($R > 0.3$) and allows a variation of about $\pm$ 1.5 Hz for a true D_a of $\sim$10 Hz.

Figure 5A indicates that exhSVD alignment tensors can deviate significantly from their true values when only a single amino acid type and $^{1}D_{NH}$ couplings are available. The problem in case of a single amino acid type and only one coupling per residue is that errors in the structure can easily be compensated by an incorrect assignment of RDCs to internuclear vectors in combination with an incorrect alignment

tensor. For example, the six RDCs observed for the six glutamines of ubiquitin can be best-fit almost perfectly to the X-ray structure, i.e., rmsd = 0.01 Hz, using exhSVD. However, five of the six RDCs are assigned to a wrong internuclear vector and the alignment tensor deviates by a generalized angle of 102° (60° for the z-axis) from the true orientation. On the other, for ILE, LEU, LYS and THR the alignment tensor is accurately determined with an average deviation of $\vartheta = 19°$ (10° for z-axis) from the true orientation.

The robustness of the method can already be significantly improved when two amino acid types are used simultaneously. Figure 5A shows results when amino acids that give incorrect alignment tensors (GLY, ARG, GLN) are combined with a second amino acid type (GLY, ARG, GLN, GLU, ILE, LYS, THR, ASP, VAL, SER, ALA, ASN, PHE, HIS, TYR). For many cases the alignment tensor determined by exhSVD is now very close to the tensor determined by SVD. Especially, whenever an amino acid is combined with either ILE, LEU, LYS or THR (each of them occurs more than seven times in ubiquitin) dynamic averaging of RDCs and deviations of the structure in solution from that in the crystal are not problematic anymore. The GLY-ILE, GLY-LYS, GLY-THR, ARG-ILE, ARG-LYS, ARG-THR, GLN-ILE, GLN-LYS, and GLN-THR exhSVD alignment tensors deviate on average by $\vartheta = 15°$ from the correct orientation (8° for the z-axis) and errors are within Monte-Carlo uncertainty estimates. The wrong alignment tensors visible in Figure 5A, are for combinations where both amino acids are problematic, for example GLY-GLN or ARG-GLY, or where the second amino acid type only has two or three residues, for example GLY-PHE or GLY-ASN. The two or three additional RDCs do not sufficiently constrain the alignment tensor. Note however, that Figure 5A just shows the worst case data where at least one amino acid type is used that is strongly affected by structural errors. There are many more amino acid combinations involving ILE, LEU, LYS, THR, VAL that are not shown in Figure 5A, but that result in highly accurate exhSVD alignment tensors. Using a combination of three amino acid types further increases the robustness of exhSVD. Figure 5A again only shows the worst case data, i.e., where at least one of the three amino acid types is strongly affected by structural errors. In most cases exhSVD alignment tensors deviate only slightly from the correct orientation and alignment tensor uncertainties are reliably estimated by the 'structural noise Monte-Carlo method'. When four amino acid types are used, the orientation of the exhSVD alignment tensor deviates only in a single case, ARG-GLY-ASP-TYR, strongly from the true orientation.

Similar results are obtained for MBP (Figure 5B). Despite MBP's size of 370 residues, five amino acid types (ARG, GLN, HIS, MET and TRP) occur less than ten times (Table 4). These amino acid types can potentially be used either separately or in combination during exhSVD. When $^1D_{NH}$ couplings of one or two amino acid types are used, exhSVD alignment tensors can deviate strongly from the true alignment tensor (Figure 5B). For a combination of three amino acid types ARG, MET and HIS are particularly suited. MBP contains six ARG, six MET and three HIS residues. This results in a total of 6!*6!*3! = 3110400 assignment permutations that have to be evaluated during exhSVD requiring less than ten minutes on a single processor SGI workstation. The alignment tensor obtained by exhSVD on the ARG-MET-HIS combination is very accurate deviating by a generalized angle of 17° from the true orientation. This corresponds to just 8° error in the orientation of the z-axis. When the 1DMB X-ray structure is used instead of the 1OMP structure, similar results are obtained. In agreement with a slightly worse SVD-fit of RDCs to 1DMB compared to 1OMP (correlation between experimental and SVD-calculated $^1D_{NH}$ RDCs are 0.92 and 0.93, respectively), the accuracy of the ARG-MET-HIS exhSVD alignment tensor is slightly reduced with $\vartheta = 24°$ and an error in z-axis orientation of 13°. In this respect it is also interesting to note that only 12 of the 15 possible $^1D_{NH}$ couplings have been reported (Yang et al., 1999) and were therefore available for exhSVD. This demonstrates that, when three different amino acid types can be identified by selective ^{15}N -labeling, it is possible to determine the molecular alignment tensor of the 370 residue protein MBP in less than 10 minutes (data analysis time) without backbone resonance assignment and not requiring ^{13}C-labeling. In addition, $^1D_{NH}$ couplings can be obtained from simple 2D HSQC spectra as spectral overlap is minimal for selectively labeled samples. Therefore, in most cases no $^1D_{NH}$ will be lost due to overlap and peak picking can be done fully automatic not requiring manual inspection of spectra. Thus, the total measurement and data analysis time that is necessary to obtain the alignment tensor of a 370 residue protein can be reduced to six hours or even less.

Rapid analysis of protein-protein interactions

For a rigid structure alignment tensors in different parts of a molecule or in different components of a molecular complex must be identical. Thus, the relative orientation of, for example two proteins in a complex, can be determined (taking into account a four-fold degeneracy) by orienting the two proteins such that their respective alignment tensors are collinear (Losonczi et al., 1999). exhSVD on the other hand, allows the determination of molecular alignment tensors without backbone resonance assignment. A strategy for rapid structure determination of protein complexes, such as the phosphoryl transfer complex between the N-terminal domain of enzyme I (EIN) and histidine-containing protein (HPR), can therefore be the following: The backbone of the smaller of the two proteins (HPR; 85 residues) will be assigned by standard triple-resonance methods. Then residues of HPR involved in binding to the bigger protein (EIN; 249 residues) are identified by chemical shift perturbation studies. In the next step $^{1}D_{NH}$ couplings in HPR will be measured and the molecular alignment tensor of HPR will be determined by a regular SVD calculation. In the standard approach the same procedure has to be done for the 249 residue protein EIN, i.e., backbone assignment, measurement of RDCs and SVD. This, however, requires $^{13}C/^{15}N$-labeled protein samples, several days of measurement time for triple-resonance experiments and another one or two weeks for the assignment process. By using exhSVD, on the other hand, backbone assignment of EIN is not required any more. Three selectively ^{15}N-labeled protein samples of EIN can be prepared and the alignment tensor of EIN can be extracted by exhSVD. This avoids ^{13}C-labeling and is therefore significantly cheaper, it requires much less measurement time as only two-dimensional ^{15}N-^{1}H HSQC are required, and data analysis can be performed fully automatic in a few minutes. After the relative orientation of HPR and EIN is established from the SVD tensor of HPR and the exhSVD tensor of EIN, the two proteins can be docked using chemical shift perturbations observed for HPR (McCoy and Wyss, 2002). Thus, the assignment of the 249 residue protein EIN is neither for determination of the relative orientation nor for docking of the two proteins required. The assignment of both EIN and HPR can potentially be avoided altogether, if it is possible to identify the interaction surface by biochemical studies or by measurement of amide proton/deuterium exchange using mass spectroscopy (Mandell et al., 1998).

Similarly, the determination of protein-ligand complexes can be accelerated significantly. For example, determination of the structure of a-methyl mannose bound to 53 kDa mannose-binding protein-A (MBPA) relied on the fact that the MBPA mutant used by Bolon et al. is trimeric in solution (Bolon et al., 1999). Due to the 3-fold symmetry the alignment tensor of MBPA has to be axially symmetric and along the symmetry axis of the molecule obviating the need for backbone assignment of MBPA. Using exhaustive back-calculation this approach can now be extended to non-symmetric molecules.

Here, we demonstrate a slightly different application of exhSVD: Rapid discrimination between monomeric and homomultimeric proteins in solution. Our test case is the 89 residue barrier-to-autointegration factor (BAF). According to its 1.9 Å crystal structure BAF is homodimeric. The homodimerization, however, can potentially be caused by crystallization and the question is if BAF is homodimeric in solution, too. In principle, there are many different ways to identify the multimeric state of a protein in solution, such as sedimentation equilibrium experiments or size-exclusion chromatography. In addition, NMR tools such as relaxation or diffusion measurements can be performed. Very often, however, these experiments do not give consistent results. On the other hand, we recently demonstrated that the alignment tensor of a protein dissolved in nearly neutral bicelles can be predicted accurately from its three-dimensional shape (Zweckstetter and Bax, 2000). As the shape of a protein differs significantly between a homodimeric assembly and a monomeric state the corresponding alignment tensors will also be very different. Therefore, by comparison of these two tensors with an experimental alignment tensor, the multimeric state of a protein in solution can be identified unambiguously (Zweckstetter and Bax, 2000).

exhSVD can accelerate this approach, as assignment of BAF is not necessary any more. This is shown in Figure 6. Experimental $^{1}D_{NH}$, $^{1}D_{NC'}$, $^{1}D_{C'CA}$ couplings have been reported previously (Cai et al., 1998), and the alignment tensor of BAF is determined without backbone resonance assignment from these RDCs using exhSVD. This alignment tensor is then compared to the alignment tensors predicted from the shape of the dimeric and the monomeric structure. Figure 6 demonstrates that alignment magnitude and orienta-

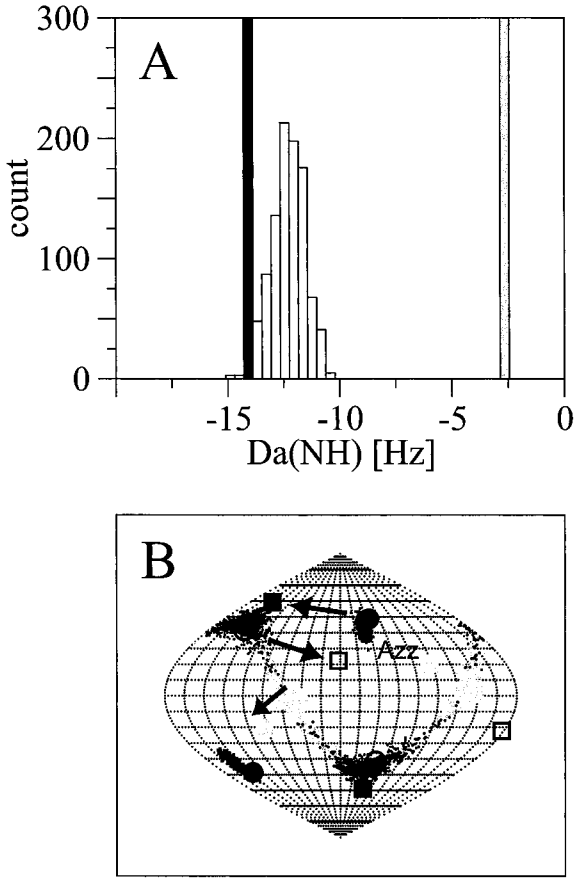

Figure 6. Comparison of the experimental alignment tensor of BAF determined by exhaustive back-calculation and alignment tensors predicted from the three-dimensional shape of homodimeric and monomeric BAF. Experimental $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C'CA}$ couplings and a 1.9 Å crystal structure (PDB code: 1CI4) were used. (A) Alignment magnitude: exhaustive back-calculation (white); homodimeric shape-prediction (black); monomeric shape-prediction (gray). (B) Alignment tensor orientation (z-axis/black with label Azz; y-axis/gray; x-axis/black): Dots (exhaustive back-calculation), ● (homodimeric shape-prediction), ■ (monomeric shape-prediction). The change in orientation when going from the homodimeric to the monomeric structure is indicated by arrows.

tion obtained by exhSVD on BAF's nine glycines are only consistent with a homodimeric assembly of BAF. This is in agreement with the NMR solution structure of BAF (Cai et al., 1998). Thus, the multimeric state of a protein can be validated unambiguously without backbone resonance assignment using a combination of shape-prediction and exhSVD. Note that RDCs were used as they have been deposited in the Protein Data Bank. No attempt was made to identify potential measurement errors, although $^1D_{C'CA}$ couplings correlate rather poorly with the BAF's X-ray structure (correlation between experimental and SVD-calculated $^1D_{C'CA}$ couplings is only 0.87 compared to 0.97 for $^1D_{NH}$). This indicates that exhSVD is quite robust against less accurate RDC measurements.

Concluding remarks

Exhaustive back-calculation allows the determination of molecular alignment tensors of proteins up to 80 kDa without prior backbone assignment. It is based on the fact that residues belonging to a specific amino acid type can be identified in a straight forward manner by either amino acid specific labeling, by their characteristic C^α/C^β chemical shifts or by dedicated NMR experiments and that the occurrence of amino acid types in proteins is highly nonuniform. exhSVD can be applied as long as at least one amino acid type is less than about 11 times present in the primary sequence of the protein. Therefore, the actual size of the protein is not important: The approach works in the same way independent of whether the protein has a total size of 80 residues or 800 residues provided that the quality of the reference structure is the same and the accuracy of experimental RDCs is comparable. No backbone assignment is required in order to extract the important information encoded by residual dipolar couplings. Therefore, exhSVD offers the opportunity to greatly accelerate determination of three-dimensional structures of protein-protein and protein-ligand complexes and validation of the multimeric state of a protein.

For very high resolution structures only $^1D_{NH}$ couplings of a single amino acid type measured in one alignment medium are in principal sufficient. However, due to structural errors and dynamical averaging of RDCs in solution, additional information is required. These can be RDCs between other atoms, e.g., $^1D_{NC'}$, $^1D_{C'CA}$ couplings, selective labeling of more than one amino acid type or $^1D_{NH}$ couplings from a second alignment medium. As it is not known a priori if the structure in solution differs significantly from the crystal structure and if the backbone remains unchanged upon complex formation error evaluation is essential. This can reliably be achieved by 'structural noise Monte-Carlo methods' in combination with exhSVD on more than one amino acid type.

Combination of two, three or four amino acid types during exhSVD promises to be particularly useful. It allows cross-validation of alignment tensors and greater robustness against structural errors. Moreover,

only $^1D_{NH}$ couplings are required and the expensive ^{13}C-labeling can be avoided. Even more important, however, is that $^1D_{NH}$ couplings can be measured for proteins up to 80 kDa with reasonable accuracy whereas measurement of the smaller $^1D_{NC'}$, $^1D_{C'CA}$ couplings becomes increasingly difficult above 40 kDa. In addition, $^1D_{NH}$ couplings can be measured from simple 2D HSQC spectra in 10 min from 50 μM protein samples using a cryoprobe (Hajduk et al., 1999). Therefore, exhSVD not only accelerates analysis of protein-protein or protein-ligand interactions. Exhaustive back-calculation also gives access to orientational information for large proteins that aggregate at concentrations above 50 μM, i.e., experimental conditions where 3D triple resonance experiments are not sensitive enough to allow backbone resonance assignment.

Acknowledgements

We thank Christian Griesinger for useful discussions. RDCs in MBP were kindly provided by Lewis Kay. M.Z. is funded by the DFG Emmy Noether-research programme.

References

Al-Hashimi, H.M., Gorin, A., Majumdar, A., Gosser, Y. and Patel, D.J. (2002) *J. Mol. Biol.*, **318**, 637–649.

Bolon, P.J., Al-Hashimi, H.M. and Prestegard, J.H. (1999) *J. Mol. Biol.*, **293**, 107–115.

BothnerBy, A.A. (1996) In *Encyclopedia of Nuclear Magnetic Resonance*, Grant, D.M. and Harris, R.K. (Eds.), Wiley, Chichester, pp. 2932–2938.

Bugayevskiy, L.M. and Snyder, J.P. (1995) *Map Projections: A Reference Manual*.

Cai, M., Huang, Y., Zheng, R., Wei, S.Q., Ghirlando, R., Lee, M.S., Craigie, R., Gronenborn, A.M. and Clore, G.M. (1998) *Nat. Struct. Biol.*, **5**, 903–909.

Clore, G.M. (2000) *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 9021–9025.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998) *J. Magn. Reson.*, **133**, 216–221.

Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Dotsch, V., Oswald, R.E. and Wagner, G. (1996) *J. Magn. Reson. Ser. B*, **110**, 107–111.

Emsley, J.W. (1996) In *Encyclopedia of Nuclear Magnetic Resonance*, Grant, D.M. and Harris, R.K. (Eds.), Wiley, Chichester, pp. 2788–2799.

Fischer, M.W.F., Losonczi, J.A., Weaver, J.L. and Prestegard, J.H. (1999) *Biochemistry*, **38**, 9013–9022.

Gardner, K.H., Zhang, X.C., Gehring, K. and Kay, L.E. (1998) *J. Am. Chem. Soc.*, **120**, 11738–11748.

Gayathri, C., BothnerBy, A.A., Vanzijl, P.C.M. and Maclean, C. (1982) *Chem. Phys. Lett.*, **87**, 192–196.

Gronenborn, A.M. and Clore, G.M. (1996) *Protein Sci.*, **5**, 981–981.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.

Hajduk, P.J., Gerfin, T., Boehlen, J.M., Haberli, M., Marek, D. and Fesik, S.W. (1999) *J. Med. Chem.*, **42**, 2315–2317.

Howard, B.R., Endrizzi, J.A. and Remington, S.J. (2000) *Biochemistry*, **39**, 3156–3168.

Hus, J.C., Prompers, J.J. and Bruschweiler, R. (2002) *J. Magn. Reson.*, **157**, 119–123.

Kontaxis, G., Clore, G.M. and Bax, A. (2000) *J. Magn. Reson.*, **143**, 184–196.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–&.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C. et al. (2001) *Nature*, **409**, 860–921.

Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.

Mandell, J.G., Falick, A.M. and Komives, E.A. (1998) *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14705–14710.

McCoy, M.A. and Wyss, D.F. (2002) *J. Am. Chem. Soc.*, **124**, 2104–2105.

Meiler, J., Prompers, J.J., Peti, W., Griesinger, C. and Bruschweiler, R. (2001) *J. Am. Chem. Soc.*, **123**, 6098–6107.

Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 11561–11562.

Moseley, H.N.B. and Montelione, G.T. (1999) *Curr. Opin. Struct. Biol.*, **9**, 635–642.

Mueller, G.A., Choy, W.Y., Yang, D.W., Forman-Kay, J.D., Venters, R.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **300**, 197–212.

Ou, H.D., Lai, H.C., Serber, Z. and Dotsch, V. (2001) *J. Biomol. NMR*, **21**, 269–273.

Peti, W., Meiler, J., Bruschweiler, R. and Griesinger, C. (2002) *J. Am. Chem. Soc.*, **124**, 5822–5833.

Ramirez, B.E. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 9106–9107.

Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2001) *J. Biomol. NMR*, **21**, 275–280.

Saupe, A. and Englert, G. (1963) *Phys. Rev. Lett.*, **11**, 462–464.

Schubert, M., Oschkinat, H. and Schmieder, P. (2001) *J. Magn. Reson.*, **153**, 186–192.

Sharff, A.J., Rodseth, L.E., Spurlino, J.C. and Quiocho, F.A. (1992) *Biochemistry*, **31**, 10657–10663.

Skrynnikov, N.R., Goto, N.K., Yang, D.W., Choy, W.Y., Tolman, J.R., Mueller, G.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **295**, 1265–1273.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Tian, F., Valafar, H. and Prestegard, J.H. (2001) *J. Am. Chem. Soc.*, **123**, 11791–11796.

Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1697–1697.

Tjandra, N., Feller, S.E., Pastor, R.W. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 12562–12566.

Tjandra, N., Grzesiek, S. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 6264–6272.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.

Tolman, J.R. (2002) *J. Am. Chem. Soc.*, **124**, 12020–12030.

Tolman, J.R., Al-Hashimi, H.M., Kay, L.E. and Prestegard, J.H. (2001) *J. Am. Chem. Soc.*, **123**, 1416–1424.

Tugarinov, V., Muhandiram, R., Ayed, A. and Kay, L.E. (2002) *J. Am. Chem. Soc.*, **124**, 10025–10035.

Umland, T.C., Wei, S.Q., Craigie, R. and Davies, D.R. (2000) *Biochemistry*, **39**, 9130–9138.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J. et al. (2001) *Science*, **291**, 1304–1351.

Vermeulen, A., Zhou, H.J. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 9638–9647.
Vijaykumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.
Warren, J.J. and Moore, P.B. (2001) *J. Magn. Reson.*, **149**, 271–275.
Yang, D.W., Venters, R.A., Mueller, G.A., Choy, W.Y. and Kay, L.E. (1999) *J. Biomol. NMR*, **14**, 333–343.
Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.
Zweckstetter, M. and Bax, A. (2001) *J. Am. Chem. Soc.*, **123**, 9490–9491.
Zweckstetter, M. and Bax, A. (2002) *J. Biomol. NMR*, **23**, 127–137.